



## A frequency-domain derivation of shot-noise

Frank Rice

Citation: *American Journal of Physics* **84**, 44 (2016); doi: 10.1119/1.4934706

View online: <http://dx.doi.org/10.1119/1.4934706>

View Table of Contents: <http://scitation.aip.org/content/aapt/journal/ajp/84/1?ver=pdfcov>

Published by the [American Association of Physics Teachers](#)

### Articles you may be interested in

[A random walk to stochastic diffusion through spreadsheet analysis](#)

*Am. J. Phys.* **81**, 823 (2013); 10.1119/1.4818410

[Unified derivation of Johnson and shot noise expressions](#)

*Am. J. Phys.* **74**, 438 (2006); 10.1119/1.2174034

[A shot-noise experiment with computer control and data acquisition](#)

*Am. J. Phys.* **73**, 984 (2005); 10.1119/1.1881272

[Shot Noise as a Tool to Probe Microscopic Interactions](#)

*AIP Conf. Proc.* **665**, 584 (2003); 10.1063/1.1584936

[Radio-frequency single-electron transistor: Toward the shot-noise limit](#)

*Appl. Phys. Lett.* **79**, 4031 (2001); 10.1063/1.1424477



American Association of **Physics Teachers**

Explore the **AAPT Career Center** –  
access hundreds of physics education and  
other STEM teaching jobs at two-year and  
four-year colleges and universities.

<http://jobs.aapt.org>



# A frequency-domain derivation of shot-noise

Frank Rice<sup>a)</sup>

California Institute of Technology, 103-33, Pasadena, California 91125

(Received 10 June 2015; accepted 15 October 2015)

A formula for shot-noise is derived in the frequency-domain. The derivation is complete and reasonably rigorous while being appropriate for undergraduate students; it models a sequence of random pulses using Fourier sine and cosine series, and requires some basic statistical concepts. The text here may serve as a pedagogic introduction to the spectral analysis of random processes and may prove useful to introduce students to the logic behind stochastic problems. The concepts of noise power spectral density and equivalent noise bandwidth are introduced.

© 2016 American Association of Physics Teachers.

[<http://dx.doi.org/10.1119/1.4934706>]

## I. INTRODUCTION

The staccato sound of raindrops falling on a tin roof, the impulses of momentum transferred by individual molecules of a gas hitting a wall of its container, the current fluctuations due to the random, independent arrivals of individual electrons flowing through certain electronic devices: in each of these examples, there may be a long-term mean rate of arrival of the pulses, but to a very good approximation the occurrence time of each individual pulse event may be modeled as completely random and independent of the times of the others. Consequently, the number of pulses occurring in equal time intervals fluctuates randomly about its average, and the interval between successive pulses varies widely, as illustrated in Fig. 1.

Because of the random variation in the time interval between successive pulses, the observed average pulse rate will show ever larger fluctuations as the averaging time interval  $\Delta t$  gets shorter, as shown in Fig. 2. These fluctuations in the observed rate of a random sequence of pulses were mathematically analyzed in 1909 by the British physicist Norman R. Campbell while considering the signals produced by radiation detectors.<sup>1</sup> In 1918, the German physicist Walter Schottky, while working in a research lab of the company now named *Siemens AG*, showed that similar fluctuations generate what he called *shot noise* in vacuum tube amplifiers.<sup>2</sup>

The event times of such independent, random pulses may be modeled as samples of a *random variable* with a uniform probability distribution. The expected value of the number of observed pulses during any particular time interval of duration  $\Delta t$  is thus  $r\Delta t$ , where  $r$  is the long-term mean pulse rate of the random sequence. The variance in the observed average pulse rate (mean squared fluctuations about  $r$ ) will vary as  $r/\Delta t$ , as illustrated in Fig. 2 and derived in Subsection 1 of the [Appendix](#). As we shall see, this fact implies that these

rate fluctuations produce *white noise* in a measuring system monitoring the sequence.

If the individual pulses are very short compared to both their mean separation and the instrumentation's best time resolution, then the detailed time structure of an individual pulse is irrelevant, and only the integrated area of a pulse will matter. In other words, the effect of the pulse is to provide an *impulse* to the system. For example, if it is a short pulse of current in a circuit, then its effect is to inject a charge (the integral of the current over the duration of the pulse) into the circuit. Continuing with this example of impulses of charge, the effect of the random pulses would be to produce a current that fluctuates about some mean value similar to the wave forms shown in Fig. 2. (The mean value would be the dc current read by a meter.)

A measuring system might be used to process a signal from just such a long sequence of random pulses, to determine two statistical quantities related to it. The first is the long-term mean of the signal (dc component, in the case of an electrical signal), and the second is the variance of the signal within some frequency range, as filtered by the system (the square of the ac RMS amplitude of the fluctuations in the measuring system output). As will be seen, these two quantities are determined by the impulse size and mean pulse rate of the random sequence, and analysis of a suitable measurement can provide estimates of these fundamental parameters.

The most straightforward approach to the determination of the relationship between a measuring system's output, with its limited frequency response, and the random pulse input stream is to calculate the frequency spectrum of that input.<sup>3</sup> This frequency-domain representation can provide the average *power spectral density* expected of the pulse sequence,

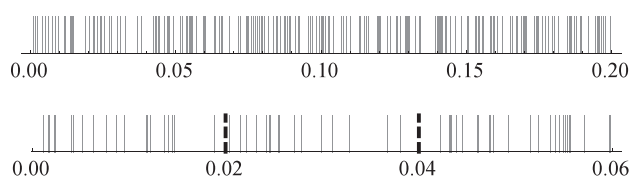


Fig. 1. A 0.2-s portion of a simulated time sequence of random, independent pulses with a long-term mean rate of  $1000 \text{ s}^{-1}$ . The first 0.06 s of the sequence is detailed in the lower plot. The actual number of pulses in the upper plot is 219, and the three 0.02-s portions of the lower plot contain 10, 17, and 26 pulses.

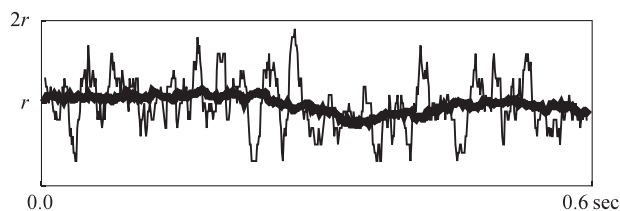


Fig. 2. Variation in the observed average pulse rate gets larger as the averaging time interval  $\Delta t$  is shortened. The data are drawn from the same set as in Fig. 1. The long-term mean pulse rate  $r$  is  $1000 \text{ s}^{-1}$ . Plotted are running means of the actual pulse rates for averaging times  $\Delta t = 0.01 \text{ s}$  and  $\Delta t = 0.1 \text{ s}$  (bold). The variance in the observed pulse rate grows as  $1/\Delta t$ , characteristic of white noise.

which may then be multiplied by the power-gain frequency response of the measurement system. The result will be the expected output signal power spectral density. To begin this analysis, in Sec. II, we derive a Fourier series representation of a typical random sequence of independent, narrow pulses.

## II. A FOURIER REPRESENTATION OF A PULSE SEQUENCE

Recall the general Fourier series representation of a periodic function  $y(t)$  with period  $T$  is

$$y(t) = y_0 + \sum_{n=1}^{\infty} (a_n \cos \omega_n t + b_n \sin \omega_n t), \quad (1)$$

with  $\omega_n = 2\pi n/T$ . The mean of  $y(t)$  over the interval  $T$  is  $\langle y(t) \rangle = y_0$ . The other coefficients are given by the inner products of  $y(t)$  with the various (orthogonal) basis functions<sup>4</sup>

$$y_0 = \frac{1}{T} \int_{-T/2}^{T/2} y(t) dt, \quad (2)$$

$$a_n = \frac{2}{T} \int_{-T/2}^{T/2} y(t) \cos \omega_n t dt, \quad (3)$$

and

$$b_n = \frac{2}{T} \int_{-T/2}^{T/2} y(t) \sin \omega_n t dt. \quad (4)$$

An infinitely long, random sequence of pulses is certainly not periodic. Consider instead a very long but finite time interval  $T$  that happens to include a very large number  $N$  of the randomly distributed pulses.<sup>5</sup> This interval  $T$  may then be interpreted as one cycle of a periodic function that can be represented by a Fourier series, ultimately letting  $T \rightarrow \infty$ . The reasoning behind this particular choice of representation is outlined in Subsection 2 of the Appendix.

Further assume that the pulses are all identical and narrow. We may conveniently model a pulse with maximum duration  $\tau$  and impulse  $q$  as rectangular with width  $\tau$  and amplitude  $q/\tau$ . If  $1/\tau \gg f_{\max}$ , where  $f_{\max}$  is the maximum frequency for which the measurement system has a detectable response, then the assumption of a narrow pulse is justified.<sup>6</sup> (Subsection 4 of the Appendix provides a straightforward method to extend the results to the case where the pulses have significant widths.)

Consider the Fourier series for a single pulse centered at the origin, given by  $t_{\text{pulse}} = 0$  (repeated with period  $T$ ). The first term in Eq. (1),  $y_0$ , is simply the mean value of the function over the period  $T$ , so  $y_0 = \langle y(t) \rangle = q/T$ . The sine terms in Eq. (1) all vanish because sine is an odd function of  $t$  whereas the rectangular pulse is even. At the origin the cosine is unity, and the pulse may be considered to be non-vanishing only near this point. Consequently, each cosine integral evaluates to  $q$ , the integral of the pulse (for all  $\omega_n \ll 1/\tau$ ). The Fourier series for a single pulse at the origin is therefore

$$y_{\text{pulse}}(t) = \frac{q}{T} + \sum_{n=1}^{n_{\max}} \frac{2q}{T} \cos \omega_n t + \dots, \quad (5)$$

with  $n_{\max} = Tf_{\max}$ . All harmonic components have the same amplitude, which is twice the mean value of the single pulse over a period. The sum has been truncated to frequency  $f_{\max}$ , because that is the highest frequency the measurement system can handle. The remaining infinite series of terms for  $n > n_{\max}$  is represented by the ellipsis. Subsection 2 of the Appendix briefly discusses the behaviors of the Fourier coefficients as  $\omega_n$  approaches or exceeds  $1/\tau$ , and Subsection 3 of the Appendix provides a more thorough examination of the effects of the measuring system's frequency response.

Now consider the complete sequence of  $N$  pulses during time  $-T/2$  to  $+T/2$ , to which we randomly assign a unique integer index  $j$  to each of the pulses, where  $1 \leq j \leq N$ . Note that the value of the index  $j$  of a particular pulse should have nothing to do with the time-ordering of the pulses, so that the  $j$ th pulse is equally likely to have occurred anywhere in the interval.<sup>7</sup> The Fourier representation of the  $j$ th pulse, centered at time  $t_j$ , will then be the same as Eq. (5), except that the time  $t$  is replaced with  $(t - t_j)$ . Since the function  $y(t)$  of this complete pulse sequence is just the sum of the individual  $y_{\text{pulse}}$  functions centered at their respective times  $t_j$ , and the expressions for the coefficients in Eqs. (2)–(4) are linear in  $y(t)$ , the Fourier representation of the complete sequence of pulses during time  $T$  will be the sum of the individual pulse representations, or

$$y(t) = \frac{Nq}{T} + \frac{2q}{T} \sum_{j=1}^N \left( \sum_{n=1}^{n_{\max}} \cos \omega_n (t - t_j) \right) + \dots \quad (6)$$

Now using a trigonometric identity and rearranging the terms in the double-sum to obtain a representation like that in Eq. (1), we get

$$y(t) = y_0 + \sum_{n=1}^{n_{\max}} (a_n \cos \omega_n t + b_n \sin \omega_n t) + \dots,$$

where

$$y_0 = \frac{Nq}{T}, \quad (7)$$

$$a_n = \frac{2q}{T} \sum_{j=1}^N \cos \omega_n t_j, \quad (8)$$

and

$$b_n = \frac{2q}{T} \sum_{j=1}^N \sin \omega_n t_j. \quad (9)$$

From trial to trial, the pulse times  $t_j$  will vary randomly throughout the interval  $T$ . We note that  $N$ , the total number of pulses within  $T$ , will vary as well. Thus, the Fourier coefficients vary randomly, so more work must be done to make use of this expression.

## III. THE SHOT NOISE EQUATION

The mutual orthogonality of the Fourier basis functions on the fundamental interval  $T$  implies that for any particular pulse sequence the mean square of  $y(t)$ ,  $\langle y(t)^2 \rangle$ , is straightforward to express in terms of the coefficients of its Fourier series representation as

$$\langle y(t)^2 \rangle \equiv \frac{1}{T} \int_{-T/2}^{T/2} y(t)^2 dt = y_0^2 + \frac{1}{2} \sum_{n=1}^{\infty} (a_n^2 + b_n^2). \quad (10)$$

The above equation is a specific example of Parseval's identity;<sup>8</sup> it shows that each Fourier component of a complicated signal contributes independently to the overall power in the signal. The variance of a particular  $y(t)$  about its mean  $\langle y(t) \rangle = y_0$  is given by the sum of the harmonic terms or

$$\langle [y(t) - y_0]^2 \rangle = \langle y(t)^2 \rangle - \langle y(t) \rangle^2 = \frac{1}{2} \sum_{n=1}^{\infty} (a_n^2 + b_n^2). \quad (11)$$

In the case of the sequence of pulses whose Fourier series is given by Eq. (7), the expressions for the coefficients  $a_n$  and  $b_n$  involve random variables. There are potential subtle difficulties with the statistical analysis of any stochastic (random) process. Refer to Subsection 1 of the Appendix for a brief discussion of some of the more important of these issues for the case of the shot noise process. With this caveat in mind, the statistical mean (expected value) of Eq. (11) must now be determined.

The pulse sequence over the interval  $-T/2 \leq t \leq +T/2$ , whose Fourier transform is given by Eq. (7), consists of exactly  $N$  pulses (with  $N > 0$ ). Thus, we first consider the expected value of Eq. (11) averaged over the appropriate ensemble of sequences that include exactly  $N$  pulses in the interval from  $-T/2$  to  $+T/2$ . We must therefore calculate

$$\overline{a_n^2}^N = \frac{4q^2}{T^2} \overline{\left( \sum_{j=1}^N \cos \omega_n t_j \right)^2} \quad (12)$$

and

$$\overline{b_n^2}^N = \frac{4q^2}{T^2} \overline{\left( \sum_{j=1}^N \sin \omega_n t_j \right)^2}, \quad (13)$$

where the superscript  $N$  indicates that these are  $N$ -pulse ensemble averages. The squared sums may be expanded so that

$$\begin{aligned} \overline{\left( \sum_{j=1}^N \cos \omega_n t_j \right)^2} &= \overline{\sum_{j=1}^N \cos^2 \omega_n t_j + \sum_{\substack{j,k=1 \\ j \neq k}}^N \cos \omega_n t_j \cos \omega_n t_k}} \\ &= \sum_{j=1}^N \overline{\cos^2 \omega_n t_j} + \sum_{\substack{j,k=1 \\ j \neq k}}^N \overline{\cos \omega_n t_j \cos \omega_n t_k}, \end{aligned} \quad (14)$$

and similarly for the squared sum over the sines. The averaging over the  $N$ -pulse ensemble (represented by the overline) may be interchanged with the sums in Eq. (14) because  $N$  in this expression is not a random variable, but rather a specified parameter value and, of course, the statistical averaging operation is linear.

The times  $t_j$  of the different pulses are statistically independent of one another, and each has a uniform probability distribution within the time interval  $T$ .<sup>9</sup> The probability

density for each  $t_j$  is therefore  $1/T$ , independent of the value of  $N$  and of the times of the other pulses. Recall that functions of statistically independent random variables are also independent, so trigonometric functions involving two different index values  $j$  and  $k \neq j$  vary independently over the ensemble. The statistical mean of the product of two independent functions is just the product of their means, and each of the terms in the double sum in Eq. (14) falls into this category, so that

$$\overline{\cos \omega_n t_j \cos \omega_n t_k} = (\overline{\cos \omega_n t_j})(\overline{\cos \omega_n t_k}) \quad (j \neq k),$$

and similarly for the sines.

Because the probability density of each  $t_j$  is uniform, the expected values of the various functions of the  $t_j$  are simply given by their mean values over the interval  $T$ . Each harmonic function, of course, has an integer number of cycles within  $T$ , therefore

$$\overline{\cos \omega_n t_j} = \overline{\sin \omega_n t_j} = 0 \quad (15)$$

and

$$\overline{\cos^2 \omega_n t_j} = \overline{\sin^2 \omega_n t_j} = 1/2. \quad (16)$$

As a result, the expected values of the cross terms vanish, and consequently, the expected value of each squared sum in Eq. (12) becomes  $N \times 1/2 = N/2$ . Thus, for  $N$ -pulse sequences occupying the interval  $T$ , using Eqs. (11) and (12) we obtain

$$\begin{aligned} \overline{\langle [y(t) - y_0]^2 \rangle}^N &= \frac{1}{2} \sum_{n=1}^{\infty} (\overline{a_n^2}^N + \overline{b_n^2}^N) \\ &= \frac{1}{2} \frac{4q^2}{T^2} \sum_{n=1}^{n_{\max}} \left( \frac{N}{2} + \frac{N}{2} \right) \\ &\quad + \cdots = \frac{2q^2 N}{T^2} \sum_{n=1}^{n_{\max}} 1 + \cdots \end{aligned} \quad (17)$$

The sum over the harmonics has been left in the final expression in Eq. (17), rather than replace it with  $n_{\max}$ , because each term represents the contribution of one of the various harmonic frequencies  $f_n = \omega_n/2\pi$  from  $f_1 = 1/T$  to  $f_{\max} = n_{\max}/T$ , evenly spaced at intervals of  $1/T$ . The frequency-space density of these harmonics (in hertz) is  $1/(1/T) = T$ , so that given any frequency range  $\Delta f$  below  $f_{\max}$ , the total contribution of the harmonics in that range to the sum in Eq. (17) is  $T\Delta f$ . As  $T \rightarrow \infty$ , the harmonics become dense in frequency space, and this expression becomes exact. Thus

$$\overline{\langle [y(t) - y_0]^2 \rangle}^N = \frac{2q^2 \Delta f}{T} N, \quad (18)$$

which is simply proportional to  $N$ . Although derived for  $N > 0$ , it is clearly correct for  $N = 0$ , the case with  $y(t) \equiv 0$  during  $T$ .

This expression may now be statistically averaged over the entire ensemble of pulse sequences during  $T$ , including the random variation in  $N$ . The result is

$$\overline{\langle [y(t) - y_0]^2 \rangle} = \frac{2q^2 \Delta f}{T} \bar{N} = 2q^2 \Delta f r, \quad (19)$$



since the expected value  $\bar{N} = rT$ , where  $r$  is the long-term mean pulse rate, as discussed in Subsection 1 of the Appendix. Note that this result has no explicit dependence on the duration  $T$ . The expected value of  $y_0$ , the signal's dc component, is  $\bar{y}_0 = qr$ , so Eq. (19) becomes the shot noise equation, or

$$\overline{[y(t) - y_0]^2} = 2q\bar{y}_0\Delta f. \quad (20)$$

The expected mean-squared amplitude of the fluctuation (shot noise) in a random sequence of independent pulses is proportional to the product of the expected dc (mean) value ( $\bar{y}_0 = qr$ ), the size of the impulse in a single pulse ( $q$ ), and the effective bandwidth of the measurement system ( $\Delta f$ , in Hz). We have thus found the desired relations mentioned above, between a measuring system's dc and ac signal outputs and the average rate  $r$  and impulse size  $q$  of the pulses. Note the derivation of Eq. (20) came from a careful and straightforward analysis starting from a minimal set of assumptions about the pulse sequence: only that the narrow pulses represent uniformly random, independent events all with the same impulse.<sup>10</sup> In Sec. IV, we examine some implications and limitations of this formula.

Generally, the average power transmitted by a noise source or other signal is proportional to its mean squared amplitude, as briefly mentioned in the discussion regarding Parseval's identity, Eq. (10). Equation (20) shows that shot noise power is proportional to the dc (average) signal level  $\bar{y}_0$ . This behavior may be considered to be a defining characteristic of shot noise, differentiating it from other processes such as thermal noise.

The measured shot noise power is also proportional to the system bandwidth  $\Delta f$ . Thus the shot noise power spectrum is flat, with a constant noise power spectral density. Such a power spectrum is called *white noise*, analogous to the idea that "white light" has an intensity that is roughly independent of frequency. The concept of power spectral density is discussed further in Subsection 3 of the Appendix.

Since matter is inherently discrete at atomic and subatomic levels, transport processes should exhibit some level of fluctuation, although generally they may be tiny (see, for example, Fig. 3). Unlike the fluctuations associated with thermal noise, shot noise fluctuations could remain present even at very low temperatures because their origin is not inherently due to thermal motions. Thus, analyses of shot noise effects can provide additional important insights into the nature of the microscopic, discrete events at the core of various physical processes.

## IV. DISCUSSION

### A. Using shot noise to determine the electron charge

The shot noise relation in Eq. (20) accurately describes the magnitudes of those fluctuations resulting from discrete events of a very specific nature. As stated at the end of Sec. III, these events must (1) be completely independent of each other, and (2) have a uniform probability distribution in time. In all cases of real, physical systems neither assumption will be completely accurate, but there are some cases for which these assumptions are quite reasonable. Two examples are the thermionic emission of electrons from the surface of the heated cathode in a vacuum tube (the problem considered by Schottky) and the generation of electron-hole pairs in the

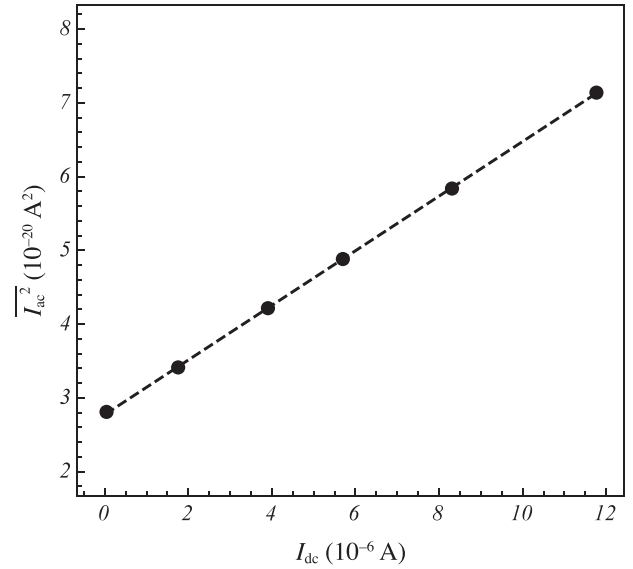


Fig. 3. Photodiode mean squared ac current vs dc current for various light source intensities. The measurement system effective bandwidth was 11.0 kHz (how this bandwidth may be determined is addressed in Subsection 3 of the Appendix). A linear fit to the data is shown, which has slope =  $(3.70 \pm 0.02) \times 10^{-15} \text{ A}$ . The measurement system's input-referred average current noise density (as inferred from the data) was  $\sim 1.6 \text{ pA}/\sqrt{\text{Hz}}$ .

depletion region of a semiconductor diode, either by thermal excitation or photoelectric absorption. In each of these cases, the phenomenon may result in an observable electric current in which the impulse size is given by the magnitude of the fundamental electron charge  $e$ .

Consider an example of the latter process (electron-hole generation). In an undergraduate laboratory experiment similar to that described by Spiegel and Helmer,<sup>11</sup> a weak light source (an infrared LED) illuminated a reverse-biased silicon photodiode. The resulting diode current was converted to a voltage by a high-gain transimpedance amplifier, and the ac noise signal of the amplifier's output was further filtered, amplified, and then squared using a precision analog multiplier circuit.<sup>12</sup> Low-pass filters served to smooth the resulting mean (dc) and variance (ac) signals, which were then measured and averaged. A typical result is shown in Fig. 3, where the system's effective bandwidth ( $\Delta f$ ) was 11.0 kHz. Using the slope of a linear fit to the data, the shot noise expression in Eq. (20) yields an estimate for the electron charge of  $1.68 \times 10^{-19} \text{ C}$ , just over 4% larger than the accepted value and within the systematic uncertainty limit determined by the measurement system's gain and bandwidth uncertainties. The average electron arrival rate  $r \approx 10^{13} \text{ sec}^{-1}$ , and the averaging time for the observations was  $\sim 1 \text{ s}$ . Clearly, the assumptions leading to Eq. (20) are not unreasonable for this particular experiment, implying that diode electron-hole pairs can be created by photon absorptions in a very nearly statistically independent manner, at least as long as the illumination is weak.

### B. Limitations of the shot noise formula

The shot noise formula in Eq. (20) accurately describes only the magnitude of fluctuations resulting from discrete events that are independent of each other. For many real transport processes, however, the assumption of statistical independence might be very poor.

For example, consider the case of electrons generated by the hot cathode of a vacuum tube diode, the problem originally considered by Schottky. A potential difference between the cathode and the diode's anode (also known as its "plate") will accelerate electrons through the vacuum separating the two electrodes, resulting in a flow of current through the diode. Electrons escape from the hot cathode in a random, independent manner much like the pulse stream shown in Fig. 1, and each starts its journey to the anode. Wherever there is a significant overdensity of electrons, however, their mutual electrical repulsion will accelerate them away from each other as their bulk motion carries them toward the anode. As a result, upon arrival at the anode, the initial electron overdensity has been moderated somewhat. This *space charge* effect can therefore reduce the magnitude of the current fluctuations through the diode, and thus reduce the level of shot noise it generates. A numerical analysis of this space charge smoothing effect is generally difficult, but shot noise in a typical vacuum tube diode's anode current may be reduced by an order of magnitude or more over Eq. (20) level.<sup>13,14</sup>

An even more striking example of shot noise suppression due to correlations among the motions of charge carriers is the remarkable uniformity of the bulk flow of electrons in a large, homogeneous conductor carrying a current induced by an applied field. In such a material, the conduction electron density is so high that Pauli exclusion keeps the electron density and drift velocity quite uniform within it, so fluctuations in the current through a cross section of the conductor are orders of magnitude smaller than that predicted by Eq. (20). This effect of quantum statistics (the conduction electrons in a metal form a type of Fermi gas) introduces strong correlations in the electrons' wave functions that nearly completely invalidate the assumption of statistical independence.<sup>15</sup> An active area of investigation is the level of shot noise suppression due to quantum and Coulomb interactions in mesoscopic electronic structures carrying "2D" or, more recently, "1D" currents.<sup>16–18</sup>

Contrary to the case of electrons, quantum correlations among photons (which are bosons) can lead to an enhancement of shot-noise-like fluctuations exceeding the level predicted by Eq. (20), a phenomenon informally known as "photon bunching." Photon bunching may be used to extract additional information about the nature of distant light sources by employing the Hanbury Brown and Twiss effect.<sup>19</sup> The fundamental nature of this phenomenon, and whether it may in some cases be mitigated to improve signal to noise ratio during observations of weak astronomical sources, is still being debated in the literature.<sup>20,21</sup>

In summary, determining the actual expected level of shot-noise-like fluctuations resulting from correlated (rather than independent) events may be quite difficult. Such studies remain an active area of research, nearly a century after Schottky's analysis.

## ACKNOWLEDGMENTS

The author gratefully acknowledges the many gracious, insightful comments and bibliographic recommendations provided by the reviewers of this manuscript. Their efforts have greatly improved its rigor and clarity.

## APPENDIX: SOME ADDITIONAL CONSIDERATIONS

### 1. The Poisson distribution and some statistical subtleties

Consider a stochastic process that may generate an event during the infinitesimal time interval  $dt$  with probability  $r dt$  ( $r$  a constant parameter), independent of the occurrences of any other past or future events generated by the process, and independent of when the interval  $dt$  may be. Such a process is termed a *Poisson process*, and the times of the events generated by the process define a set of *Poisson points*. It can be shown that the distribution of the number of events  $N$  generated by the process during a finite time interval  $T$  is given by the Poisson probability distribution<sup>22</sup>

$$P(N) = \frac{e^{-rT} (rT)^N}{N!}. \quad (\text{A1})$$

The expected value of  $N$  in Eq. (A1) is  $\bar{N} = rT$ , and so is its variance  $\sigma_N^2 = \bar{N}$ .

The Poisson process with its corresponding distribution is one of the most thoroughly analyzed stochastic processes. The text by Papoulis and Pillai,<sup>23</sup> for example, devotes many pages to it. The result  $r = \bar{N}/T$  is the mean event rate expected from the process, and the variance in the observed rate  $N/T$  around this mean would be

$$\sigma_r^2 = \frac{\sigma_N^2}{T^2} = \frac{\bar{N}}{T^2} = \frac{r}{T}. \quad (\text{A2})$$

Thus as  $T \rightarrow \infty$  the expected variance in the observed event rate  $\rightarrow 0$ , and for a long sequence of events produced by any Poisson process,  $N/T \rightarrow r$  as  $T \rightarrow \infty$ . Therefore, we may identify  $r$  with the long-term event rate. To proceed with our study of the statistical behavior of the Poisson process with a reasonable level of rigor, we must have confidence that it possesses two important properties. The first is that it is *stationary* and the second that it is *ergodic*.

A stationary process is one whose statistical properties are homogeneous in time. This means that if we choose an interval  $\Delta t$  around any arbitrary time  $t$ , then the ensemble statistics applicable to that interval are independent of when  $t$  may be. Since the infinitesimal event probability  $r dt$  described above for a general Poisson process is independent of the existence of other events in the sequence, the process is memoryless and is consequently *strict-sense stationary* as described in Ref. 23, Chapter 9.

The concept of the ergodicity of a stochastic process is quite a bit more subtle and is of vital importance to the validity of statistical calculations regarding a physical system. Consider the statistical ensemble of possible random event sequences which could be generated by an ergodic stochastic process. In this case, the expected values (statistical mean values) of properties averaged over the ensemble are equal to the limits of the corresponding time averages of a typical actual sequence, as the averaging time goes to infinity.

Ergodicity is discussed in detail in many texts on stochastic processes or statistical mechanics.<sup>23</sup> For our purposes, we assume that if time were divided into an infinite number of disjoint intervals each of duration  $T$ , then the statistics derived from comparing the actual numbers of events generated by a particular Poisson process within these various

time intervals would match that of the Poisson probability distribution given by Eq. (A1). This assumption, whose demonstration is beyond the scope of this paper, would nevertheless seem to be quite reasonable.

Taking our assumptions of stationarity and ergodicity to be valid, consider a very long time sequence of duration  $T$  and total number of events  $N$ . Equation (A2) shows that by taking  $T$  to be long enough we can make the difference between  $r$  and  $N/T$  arbitrarily small. Assume that we have done this, and the error in approximating  $r$  as  $N/T$  is negligible. Now if we were to consider the division of  $T$  into a large number of disjoint intervals  $\Delta t$ , then the distribution of the numbers of events within these subintervals will be described by Eq. (A1), and the variance around  $r$  of the event rates in the intervals will be described by Eq. (A2). Thus, as the “observation time”  $\Delta t$  is made shorter, the variance in the observed event rate rises as  $r/\Delta t$ , as mentioned in the introduction and illustrated in Fig. 2, where each pulse corresponds to an event.

The expected number of events in any subinterval of duration  $\Delta t$  in  $T$  is  $r \Delta t = N \Delta t/T$ , independent of the location of  $\Delta t$  within  $T$ . This result would also be the expected number of events in  $\Delta t$  if the  $N$  events in  $T$  were all independent and each had a uniform probability of occurrence at any time throughout  $T$ . Therefore, the assumptions of uniformity and relative independence used to derive the shot noise expression in Eq. (20) are satisfied by pulses generated by a Poisson process.

Finally, we show that the Poisson distribution of the number of events within the various intervals  $\Delta t$  is consistent with Eq. (20). If each event consists of a pulse injecting an impulse  $q$  to a measuring system, then the expected mean signal  $\bar{y}_0 = qr$ , and the expected variance around this mean would be  $q^2 \sigma_r^2$ , with  $\sigma_r^2$  given by Eq. (A2); thus

$$\langle [y(t) - y_0]^2 \rangle = q^2 \sigma_r^2 = q^2 \frac{r}{\Delta t} = \frac{q \bar{y}_0}{\Delta t}. \quad (\text{A3})$$

The required bandwidth of the measuring system in order for it to respond to the rate changes expected between successive, disjoint observations, each of duration  $\Delta t$ , is  $\Delta f \sim 1/\Delta t$ . In fact, according to the Nyquist-Shannon sampling theorem,<sup>24,25</sup>

$$\frac{1}{f} = 2 \Delta t, \quad (\text{A4})$$

because, roughly, it takes at least two samples per cycle to respond to a frequency  $f$ . Inserting this result into Eq. (A3) recovers the shot noise expression given by Eq. (20). Thus, an analysis of the Poisson statistics of the pulse times appears to be consistent with our Fourier analysis of the pulse stream, as it should be, and mean squared rate fluctuations that are inversely proportional to the averaging time  $\Delta t$  are characteristic of a white noise power spectral density.

Note that this “derivation” of the equivalence of the two approaches, frequency-domain analysis vs time-domain Poisson statistics, is not rigorous but only meant to be illustrative. A complete derivation of the shot noise power spectrum from time-domain considerations may be found in the paper by Mathieson,<sup>26</sup> while a thorough but more abstract analysis may be found in Chapter 10 of the text by Papoulis and Pillai.<sup>23</sup>

## 2. Fourier series

The shot noise problem was analyzed by using a Fourier series representation of a long but finite pulse sequence rather than by using a Fourier transform of an infinite pulse sequence. This was done for two reasons. First, the math is more straightforward with the series representation of a long but finite pulse train, because one only deals with sums of a countable number of terms rather than integrals over a continuum of Fourier components. Second, and more important, the results of a real experiment will always consist of measurements of finite quantities over a finite time interval, so the analysis presented is more representative of the actual situation. In this case, as  $T \rightarrow \infty$  the harmonic terms become dense in frequency, but remain countable.

Note also that by considering a Fourier series of the function  $y(t)$  over the interval  $T$ , it was assumed that  $y(t)$  may be represented as a periodic function (with, of course, period  $T$ ) without introducing any significant bias to the analysis. This is equivalent to applying cyclic boundary conditions to a finite (albeit large) domain of interest, a typical approach used by physicists when interested in the “bulk” properties of a macroscopic, complicated system. By doing this one might avoid, for example, the introduction of spurious artifacts to the spectral analysis that could arise when a window function is applied to a temporal and/or spatial function before taking its Fourier transform. It also has the added advantage of keeping the Fourier components (which may be associated with a system’s normal modes) countable, as mentioned above.

An interesting and instructive example of the alternate approach, the Fourier transform of a temporally windowed stochastic process, is the paper by Abbott *et al.*,<sup>27</sup> whose spectral analysis of Johnson noise may serve as an introduction to a slightly more sophisticated technique to attack problems such as this.

Another objection to the analysis may be that the mathematical manipulations carried out in Eqs. (5)–(17) were performed on a finite sum of terms rather than on the full Fourier series, including only those harmonic frequencies less than some arbitrary  $f_{\max}$ . Why not choose to consider an infinite series representing a sum of Dirac delta functions, in which case one would simply replace  $f_{\max}$  with infinity? Were this to be done, then the resulting Fourier series would have had terms all of the same magnitude, meaning the Fourier series for  $\delta(t)$  is not absolutely convergent. Thus, the series manipulations (rearranging terms and changing the order of the sums) would not be mathematically justifiable.

Actually, any real pulse should be, of course, modeled as a continuous function of the time with some finite duration characterized by a width  $\tau$ . This implies that for harmonics with  $\omega_n \tau \sim 1$ , the Fourier coefficients will start to decrease at least as fast as  $1/n$ , and for  $\omega_n \tau \gg 1$  they will fall at least as fast as  $1/n^2$ . Otherwise, for  $\omega_n \tau \ll 1$  the sum in Eq. (5) with constant  $a_n = 2q/T$  is quite accurate. The Fourier series for an actual set of observable pulses is therefore absolutely and uniformly convergent, so that the subsequent manipulations leading up to and including Eq. (20) are valid, and that result is accurate as long as the maximum frequency making a significant contribution to the system’s effective bandwidth is  $\ll 1/\tau$ . Subsections 3 and 4 of this Appendix consider a simple procedure to deal with the situation wherein the pulses have significant widths.



### 3. White noise, spectral density, and filter bandwidth

We have found that the mean squared shot noise fluctuations are expected to be proportional to the bandwidth over which they are measured, the constant of proportionality being  $2q\bar{y}_0$ . The units of this coefficient are  $(\text{amplitude})^2/\text{Hz}$ . Because the power in a signal goes as its  $(\text{amplitude})^2$ ,  $2q\bar{y}_0$  is proportional to the power spectral density of the shot noise. For example, in the case of the random flow of electrons considered in Sec. IV, the shot noise signal consists of fluctuations in electrical current. Sending this signal through a resistance  $R$  would represent a noise power spectral density of  $p(f) = 2q_e I_{dc} R W/\text{Hz}$ .

In this case, the noise power spectral density is independent of frequency (for nonzero frequencies less than  $f_{\max}$ ), and its power spectrum is called white noise. More generally, a complicated signal or one that has been passed through a filter will have a power spectral density that is a function of frequency. In this latter case, the total signal power within some specified frequency range may be calculated by integrating

$$P = \int_{f_{\min}}^{f_{\max}} p(f) df, \quad (\text{A5})$$

where  $p(f)$  is the signal's power spectral density at frequency  $f$ .

With this idea of the noise power spectral density in mind, we now turn to the problem of determining a measurement system's effective bandwidth  $\Delta f$ , for use in the shot noise expression in Eq. (20). A linear system's effect on the amplitude and phase of an input signal at frequency  $f$  is described by its generally complex-valued transfer function  $H(f)$ , which, for this elementary discussion, we take to be dimensionless.<sup>28</sup> The power gain of such a system at any particular frequency would then be  $|H(f)|^2$ . If you pass a signal with power spectral density  $p(f)$  through the system, then the signal power spectral density at the system's output will be

$$p'(f) = |H(f)|^2 p(f), \quad (\text{A6})$$

not including any additional noise the system may add to the signal. Thus, a white-noise source such as shot noise with constant power spectral density  $p$  input to the system would generate an average output power of

$$\langle P \rangle = \int_0^\infty |H(f)|^2 p df = p \int_0^\infty |H(f)|^2 df.$$

We can then define the *equivalent noise bandwidth*  $\Delta f_{\text{eq}}$  of the measurement system to be

$$\Delta f_{\text{eq}} \equiv \int_0^\infty |H(f)|^2 df. \quad (\text{A7})$$

This expression, then, is the effective bandwidth to be used in Eq. (20).<sup>29</sup> Note that our cut-off frequency  $f_{\max}$  is estimated by  $|H(f \geq f_{\max})|^2 \sim 0$ , or more precisely, replacing the upper limit of integration in Eq. (A7) with  $f_{\max}$  should have negligible effect on the resulting value of  $\Delta f_{\text{eq}}$ .<sup>30</sup> The equivalent noise bandwidths for various filter transfer functions may be calculated using Eq. (A7). Table I lists these bandwidths for a few of the most common configurations. If the power gain of a given system differs from that shown in the table, then the calculated mean squared noise amplitude

Table I. Equivalent noise bandwidths of selected filters.

Type	$ H(f) ^2$	$\Delta f_{\text{eq}}$	Comment
First order LP	$[1 + f^2/f_0^2]^{-1}$	$\frac{\pi}{2}f_0$	Pass-band gain = 1
Second order LP	$\left[1 + \left(\frac{1}{Q^2} - 2\right)\frac{f^2}{f_0^2} + \frac{f^4}{f_0^4}\right]^{-1}$	$\frac{\pi}{2}Qf_0$	Pass-band gain = 1
Second order BP	$\left[1 + Q^2\left(\frac{f^2}{f_0^2} + \frac{f_0^2}{f^2} - 2\right)\right]^{-1}$	$\frac{\pi}{2}\frac{1}{Q}f_0$	Gain at $f_0 = 1$

within  $\Delta f_{\text{eq}}$  should, of course, be multiplied by the system's power gain.

For example, a system used to estimate the electron charge from a measurement of a photodiode's shot noise intensity vs dc current (see Fig. 3) might have its bandwidth determined by a second-order Butterworth low-pass filter with a corner frequency of 10.0 kHz. Using the second entry in Table I with  $Q = 1/\sqrt{2}$  and  $f_0 = 10.0$  kHz gives an equivalent noise bandwidth of 11.1 kHz. The actual system used for the data in Fig. 3 also included a second-order high-pass filter with a corner frequency of 100 Hz, so the actual  $\Delta f_{\text{eq}} = 11.0$  kHz.

### 4. Pulses of finite width

The derivation leading up to Eq. (20) required the explicit assumption that the pulse width  $\tau$  be much shorter than the best time resolution of the measuring system ( $\sim 1/f_{\max}$ ). In many situations, this may not be the case. For example, consider the sequence shown in Fig. 4. This random sequence comprises a series of identical "tail pulses," or events with an abrupt rise followed by an exponential decay. As illustrated in the figure, these exponential tails may not be short compared to the mean inter-pulse spacing, and sequential pulses often overlap. What modification to Eq. (20) would be required for a series of random pulses such as this one?

A mathematical representation of a tail pulse starting at  $t=0$  with total impulse (integral)  $q$  and decay time constant  $\tau$  is

$$y_{\text{pulse}} = \frac{q}{\tau} e^{-t/\tau}. \quad (\text{A8})$$

It is straightforward to show that this waveform would be the result of passing a delta function impulse  $q\delta(t)$  through a first-order low-pass filter with a corner frequency of  $f_\tau = 1/2\pi\tau$ , with the complex transfer function  $H(f)$  of such a filter being ( $j \equiv \sqrt{-1}$ )

$$H(f) = \frac{1}{1 + jf/f_\tau}. \quad (\text{A9})$$

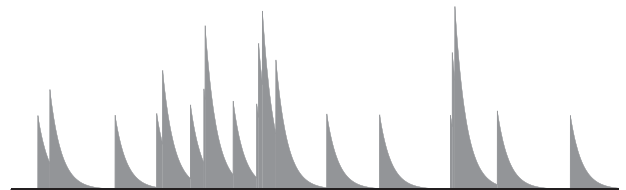


Fig. 4. In a real experiment, the pulses may have significant widths, as is the case for the random sequence of tail pulses illustrated here. Note that most of the pulses show significant overlap with their neighbors, leading to "pulse pileup."



Thus, the resulting shot noise spectral density  $y^2(f)$  generated by a random sequence of tail pulses will simply be that of Eq. (20) passed through such a filter. Using Eq. (A6) or Table I, we obtain

$$y^2(f) = \frac{2q\bar{y}_0}{1 + (f/f_\tau)^2}. \quad (\text{A10})$$

Other pulse shapes may be treated analogously.

- <sup>a)</sup>Electronic mail: rice@caltech.edu
- <sup>1</sup>Norman Campbell, “The study of discontinuous phenomena,” *Proc. Cambridge Philos. Soc.* **15**, 117–136 (1909); available at <http://www.biodiversitylibrary.org/item/97262#page/145/mode/1up>.
- <sup>2</sup>W. Schottky, “Über spontane stromschwankungen in verschiedenen elektrizitätsleitern,” *Ann. Phys. (Berlin)* **362**, since renumbered to 362, 541–567 (1918).
- <sup>3</sup>Many texts derive the shot noise expression using a purely time-domain representation (or, at least, all the heavy lifting is done in the time-domain before transforming the result to frequency space). Direct frequency-domain analysis of stochastic processes may then be subsequently introduced as a more advanced topic (Refs. 13, 23, and 31). Journal articles also tend to prefer the time-domain approach, as it is generally thought to be more flexible (Refs. 11, 26, 32, and 33). Notable exceptions are the excellent papers by Carson (Ref. 34) and by Abbott *et al.* (Ref. 27) and the out-of-print text by Goldman (Ref. 35). One must also consider the classic text by Bleaney and Bleaney (Ref. 14). Subsection 1 of the Appendix gives a very brief example of a time-domain analysis of shot noise.
- <sup>4</sup>The mean of a function  $y(t)$  of a particular pulse sequence during the time interval  $T$  will be denoted by  $\langle y(t) \rangle$ . The statistical mean of some quantity averaged over an ensemble of possible pulse sequences will be indicated by an overline:  $\overline{a_n^2}$ . Both may be applied to a function:  $\langle y(t) \rangle$  would be an ensemble average over the distribution of possible time averages  $\langle y(t) \rangle$ .
- <sup>5</sup>Unless the long-term mean pulse rate  $r$  is infinitesimally small,  $T$  can be chosen large enough for  $N$  to exceed any given finite value, although how large that  $T$  must be will be determined by the particular random-pulse sequence. For a very large minimum value for  $N$ , however, choosing  $T$  to be, say,  $2N_{\min}/r$  should ensure that in all but a very small number of “perverse” sequences the actual number  $N$  of pulses contained in  $T$  will exceed  $N_{\min}$ . The number of pulses should be approximately  $2N_{\min}$  (see Subsection 1 of the Appendix).
- <sup>6</sup>In the limit that the pulse width  $\tau$  becomes infinitesimal, a pulse with integral  $q$  centered at time  $t_{\text{pulse}}$  approaches  $q \times \delta(t - t_{\text{pulse}})$ , where  $\delta$  is the Dirac delta function. Subsection 2 of the Appendix further discusses this issue.
- <sup>7</sup>The indices  $j$  for the pulse times of a particular pulse sequence could be assigned, for example, by recording the various times on slips of paper and then drawing them from a hat. The time  $t_j$  would be that recorded on the  $j$ th slip drawn.
- <sup>8</sup>Marc-Antoine Parseval, “Mémoire sur les séries et sur l’intégration complète d’une équation aux différences partielle linéaire du second ordre, a coefficients constans,” *Sci. Math. Phys.* **1**, 638–648 (1806) available at <http://gallica.bnf.fr/ark:/12148/bpt6k3305v/f10.item>.
- <sup>9</sup>Mutual statistical independence and uniform distribution within  $T$  of the various  $t_j$  are the reasons why we required that the enumeration  $j$  of the pulses be unrelated to their time ordering.
- <sup>10</sup>The present derivation, although arrived at independently, is essentially the same as that presented in the Bleaney’s text (Ref. 14), except that we have been more careful and explicit regarding the statistical calculations.
- <sup>11</sup>D. R. Spiegel and R. J. Helmer, “Shot-noise measurements of the electron charge: An undergraduate experiment,” *Am. J. Phys.* **63**(6), 554–560 (1995).
- <sup>12</sup>The apparatus used for the example experiment was manufactured by Teachspin, Inc., Buffalo, NY. See <[www.teachspin.com](http://www.teachspin.com)>.
- <sup>13</sup>Wilbur Davenport and William Root, *An Introduction to the Theory of Random Signals and Noise* (Wiley-IEEE Press, New York, 1987).
- <sup>14</sup>B. I. Bleaney and B. Bleaney, *Electricity and Magnetism*, 1st ed. (Oxford U.P., London, 1957), reissue of the 3rd edition, in two volumes, 2013.
- <sup>15</sup>Markus Büttiker, “Flux-sensitive effects in normal metal loops,” *Ann. NY Acad. Sci.* **480**(1), 194–209 (1986).
- <sup>16</sup>Ya. M. Blanter and M. Büttiker, “Shot noise in mesoscopic conductors,” *Phys. Rep.* **336**, 1–166 (2000).
- <sup>17</sup>R. Danneau, F. Wu, M. F. Craciun, S. Russo, M. Y. Tomi, J. Salmilehto, A. F. Morpurgo, and P. J. Hakonen, “Shot noise measurements in graphene,” *Solid State Commun.* **149**, 1050–1055 (2009).
- <sup>18</sup>A. Betti, G. Fiori, and G. Iannaccone, “Enhanced shot noise in carbon nanotube FETs due to electron-hole interaction,” in *Proceedings of 14th International Workshop on Computational Electronics (IWCE)* (2010), pp. 1–4.
- <sup>19</sup>E. M. Purcell, “The question of correlation between photons in coherent light rays,” *Nature* **178**, 1449–1450 (1956).
- <sup>20</sup>Jonas Zmuidzinas, “Thermal noise and correlations in photon detection,” *Appl. Opt.* **42**, 4989–5008 (2003).
- <sup>21</sup>Richard Lieu, T. W. B. Kibble, and Lingze Duan, “A method to improve the sensitivity of radio telescopes,” *Astrophys. J.* **798**, 67–72 (2015).
- <sup>22</sup>The eponymous French mathematician Siméon Poisson published his distribution in 1837. Many texts on probability and statistics derive its relation to our uniform, random sequence of pulses; see, for example, Philip Bevington and D. Keith Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, 3rd ed. (McGraw-Hill, New York, 2003), pp. 23–27.
- <sup>23</sup>Athanasios Papoulis and S. Unnikrishna Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. (McGraw-Hill, New York, 2002).
- <sup>24</sup>H. Nyquist, “Certain topics in telegraph transmission theory,” *Proc. IEEE* **90**(2), 282–305 (2002). Originally published in *Trans. AIEE* **47**, 617–644 (1928).
- <sup>25</sup>Claude E. Shannon, “Communication in the presence of noise,” *Proc. IEEE* **86**(2), 447–457 (1998). Originally published in *Proc. IRE* **37**(1), 10–21 (1949).
- <sup>26</sup>E. Mathieson, “Derivation of noise formulas using Campbell’s theorem,” *Am. J. Phys.* **45**(6), 1184–1186 (1977).
- <sup>27</sup>Derek Abbott, Bruce R. Davis, Nicholas J. Phillips, and Kamran Eshraghian, “Simple derivation of the thermal noise formula using window-limited Fourier transforms and other conundrums,” *IEEE Trans. Educ.* **39**(1), 1–13 (1996).
- <sup>28</sup>The transfer function  $H(f)$  will not be dimensionless if the system output has units different from those of its input. For example, the input may be an electrical current whereas the output may be a voltage, in which case  $H(f)$  has units of resistance (called the system’s *transimpedance*).
- <sup>29</sup>Usually, the equivalent noise bandwidth calculated using Eq. (A7) is normalized by dividing by the power gain  $G$  of the system at some frequency  $f_0$  well within its bandwidth for normal operation:  $G = |H(f_0)|^2$ . The output mean squared signal amplitude is then finally determined by multiplying the white-noise spectral density by both the system’s equivalent noise bandwidth and its nominal power gain  $G$ .
- <sup>30</sup>Note that the integrals in Eq. (A7) and the one above it should generally not include  $f=0$  (the dc signal component) in their evaluation without some special precautions; these expressions are strictly correct only for the cases where  $|H(f \rightarrow 0)|^2 = 0$ .
- <sup>31</sup>Aldert Van der Ziel, *Noise* (Prentice-Hall, Englewood Cliffs, NJ, 1954).
- <sup>32</sup>Luca Callegaro, “Unified derivation of Johnson and shot noise expressions,” *Am. J. Phys.* **74**(6), 438–440 (2006).
- <sup>33</sup>T. Lukes, “The statistical properties of sequences of stochastic pulses,” *Proc. Phys. Soc.* **78**, 153–168 (1961).
- <sup>34</sup>J. R. Carson, “The statistical energy-frequency spectrum of random disturbances,” *Bell Syst. Tech. J.* **10**, 374–381 (1931).
- <sup>35</sup>Stanford Goldman, *Frequency Analysis, Modulation, and Noise* (McGraw-Hill, New York, 1948).